Leveraging Machine Learning and Natural Language Processing for Monitoring E-health Publications

Andrius Budrionis, Rune Pedersen, Torbjørn Torsvik, Karianne Lind, Omid Saadatfard
Norwegian Centre for E-health Research
University Hospital of North Norway

Presenter: Andrius Budrionis, Andrius.Budrionis@ehealthresearch.no
About the presenter

Andrius Budrionis, PhD
Senior Researcher, Health data and analytics
Norwegian Centre for E-health Research, University Hospital of North Norway

Background in computer science and software engineering
8+ years experience in e-health

Major focus – clinical data reuse and advanced data analytics. How to deploy and scale machine learning models?
E-health in Norway

Helsedirektoratet

Direktoratet for e-helse

Norwegian Centre for E-health Research
1. Digitization of work processes
2. Better continuity of care
3. Better use of health data
4. New ways to provide healthcare
5. Common foundation for digital services
6. National e-health management and increased implementation
Method

<table>
<thead>
<tr>
<th>Dataset</th>
<th># e-health</th>
<th># not-e-health</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-health publication dataset</td>
<td>816</td>
<td>1075</td>
</tr>
<tr>
<td>PubMed dataset</td>
<td>25</td>
<td>899</td>
</tr>
</tbody>
</table>

Publishing dates: 01-01-2010 – 01-04-2020

Author affiliation = “Norway”

N = 70058
N = 1300

Dataset

Publication dataset

E-health

# e-health

# not-e-health

816
1075

25
899
## Results. Classification

### BERT, 2-class classifier

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>f-1 score</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-health publication dataset</td>
<td>Not e-health</td>
<td>0.92</td>
<td>0.88</td>
<td>0.9</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>E-health</td>
<td>0.85</td>
<td>0.90</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>PubMed dataset</td>
<td>Not e-health</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>E-health</td>
<td>0.82</td>
<td>0.72</td>
<td>0.77</td>
<td></td>
</tr>
</tbody>
</table>

### Naive Bayes, 6-class classifier

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>f1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Digitization of work processes</td>
<td>0.70</td>
<td>0.58</td>
<td>0.63</td>
</tr>
<tr>
<td>2. Better continuity of care</td>
<td>0.61</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>3. Better use of health data</td>
<td>0.62</td>
<td>0.71</td>
<td>0.67</td>
</tr>
<tr>
<td>4. New ways to provide healthcare</td>
<td>0.74</td>
<td>0.77</td>
<td>0.75</td>
</tr>
<tr>
<td>5. Common foundation for digital services</td>
<td>0.53</td>
<td>0.62</td>
<td>0.57</td>
</tr>
<tr>
<td>6. National e-health management and increased implementation</td>
<td>0.66</td>
<td>0.64</td>
<td>0.65</td>
</tr>
</tbody>
</table>
Results. Distribution of e-health publication during last 10 years
Results. Classification

Top-5 keywords representing each class (sorted by importance):

1. **Digitization of work processes**: nurse, patient, use, care, hospital
2. **Better continuity of care**: patient, care, inform, health, communication
3. **Better use of health data**: data, patient, health, record, predict
4. **New ways to provide healthcare**: diabetes, patient, health, use, social
5. **Common foundation for digital services**: secure, health, standard, information, develop
6. **National e-health management and increased implementation**: telemedicine, health, information, implement, studies
Results. Classified e-health publication stratified yearly
Discussion

• Limitations of data collection
• Classifier performance
• Alternative classification strategies
Way forward

• Integration of other data sources (WoS, Scopus, etc...)
• Use of full-text?
• More granular interactive visualizations